



# UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE  
United States Patent and Trademark Office  
Address: COMMISSIONER FOR PATENTS  
P.O. Box 1450  
Alexandria, Virginia 22313-1450  
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/684,773	10/14/2003	Zicheng Liu	M61.12-0547	3356
27366	7590	09/07/2006	EXAMINER	
WESTMAN CHAMPLIN (MICROSOFT CORPORATION) SUITE 1400 900 SECOND AVENUE SOUTH MINNEAPOLIS, MN 55402-3319			WOODS, ERIC V	
			ART UNIT	PAPER NUMBER
			2628	

DATE MAILED: 09/07/2006

Please find below and/or attached an Office communication concerning this application or proceeding.

<b>Office Action Summary</b>	Application No.	Applicant(s)
	10/684,773	LIU ET AL
Examiner	Art Unit	
Eric Woods	2628	

*-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --*

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

## Status

1)  Responsive to communication(s) filed on 15 June 2006.

2a)  This action is FINAL.                            2b)  This action is non-final.

3)  Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

## **Disposition of Claims**

4)  Claim(s) 1,2,6,7,10,13-29,33-39 and 43-47 is/are pending in the application.  
4a) Of the above claim(s) \_\_\_\_\_ is/are withdrawn from consideration.

5)  Claim(s) \_\_\_\_\_ is/are allowed.

6)  Claim(s) 1-2,6-7,10,13-29,33-39,43-47 is/are rejected.

7)  Claim(s) \_\_\_\_\_ is/are objected to.

8)  Claim(s) \_\_\_\_\_ are subject to restriction and/or election requirement.

## Application Papers

9)  The specification is objected to by the Examiner.

10)  The drawing(s) filed on \_\_\_\_\_ is/are: a)  accepted or b)  objected to by the Examiner.

Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

11)  The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

**Priority under 35 U.S.C. § 119**

12)  Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).  
a)  All    b)  Some \* c)  None of:  
1.  Certified copies of the priority documents have been received.  
2.  Certified copies of the priority documents have been received in Application No. \_\_\_\_\_.  
3.  Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

\* See the attached detailed Office action for a list of the certified copies not received.

**Attachment(s)**

1)  Notice of References Cited (PTO-892) 4)  Interview Summary (PTO-413)  
2)  Notice of Draftsperson's Patent Drawing Review (PTO-948) Paper No(s)/Mail Date. \_\_\_\_ .  
3)  Information Disclosure Statement(s) (PTO/SB/08)  
Paper No(s)/Mail Date \_\_\_\_ . 5)  Notice of Informal Patent Application  
6)  Other: \_\_\_\_ .

## DETAILED ACTION

### ***Continued Examination Under 37 CFR 1.114***

A request for continued examination under 37 CFR 1.114, including the fee set forth in 37 CFR 1.17(e), was filed in this application after final rejection. Since this application is eligible for continued examination under 37 CFR 1.114, and the fee set forth in 37 CFR 1.17(e) has been timely paid, the finality of the previous Office action has been withdrawn pursuant to 37 CFR 1.114. Applicant's submission filed on 6/15/2006 has been entered.

### ***Response to Arguments***

Applicant's arguments, see Remarks pages 1-3 and claim amendments, filed with the RCE on 6/15/2006, with respect to the rejection(s) of claim(s) 1-43 under various statutes have been fully considered and are persuasive.

Claims 3-5, 8-9, 11-12, 30-32, and 40-42 have been canceled.

Claims 44-47 have been added.

Therefore, in view of applicant's amendments to the independent claims 1, 25, and 35, the rejection of all remaining claims (the others are dependent upon those independent claims) has been withdrawn.

It is respectfully noted that applicants state on page 1 of Remarks that the independent claims were rejected under 35 USC 102 in the last action. However, an examination of the record shows that the only grounds of rejection used in the Final Action issued on 1/12/2006 were made under 35 USC 103.

However, upon further consideration, a new ground(s) of rejection is made in view of various references as below.

Examiner would further point out that there is no statutory or regulatory basis or anything in the MPEP that states that examiner is required to consider the preamble.

Further, the CAFC has repeatedly stated (see *Scanner Technologies v. IKOS Vision Systems* – CAFC 2004) that the singular article adjective 'a' or 'an' is properly construed as meaning 'one or more'. Therefore, applicant's arguments that the claims are directed to modifying a single frame are not found to be persuasive and are not being given patentable weight, because they are properly read as 'one or more single frames,' and further the recitations occur in the preamble (see *Pitney Bowes v. Hewlett-Packard Co.*, 182 F.3d 1298, 1305, 51 USPQ2d 1161, 1165 (Fed. Cir. 1999)).

Next, examiner will present the following as a pre-emptive defense against the inevitable hindsight argument. Applicant freely admits in the specification that their work is loosely based upon or derived from Pighin et al, and explores it in some depth – instant specification pages 12-13. At the end of the Pighin et al paper, several improvements are suggested, and branches for future work. Therefore, it is **not** improper or impermissible hindsight to ask the question: would taking the work of Pighin et al and including some of the suggested future work render the claimed invention of applicant obvious? Examiner maintains that there can be no hindsight when a primary reference (Pighin) expressly suggests making certain improvements; those improvements (and their attendant modifications and teachings) are completely and totally obvious, with the modification inherent within the primary reference (e.g.

Future work section of Pighin). The *Graham* tests – and clearly the Federal Circuit's TSM test – are both met by statements in the Pighin reference itself.

### ***Claim Objections***

Claims 1, 25, and 35 are objected to because of the following informalities: the series of wherein clauses in the second portion of the claim is not appropriately listed; that is, they should be separated by semicolons, not a string of them that appears as: A and B and C and so forth. Appropriate correction is required.

Claim 46 is objected to because the last line contains the word "liner" which should obviously be "linear".

### ***Claim Rejections - 35 USC § 112***

The following is a quotation of the second paragraph of 35 U.S.C. 112:

The specification shall conclude with one or more claims particularly pointing out and distinctly claiming the subject matter which the applicant regards as his invention.

Claims 6, 33, and 40 are rejected under 35 U.S.C. 112, second paragraph, as being indefinite for failing to particularly point out and distinctly claim the subject matter which applicant regards as the invention.

Claims 6, 33, and 40 are dependent upon a canceled claim.

### ***Claim Rejections - 35 USC § 103***

The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

(a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negatived by the manner in which the invention was made.

The factual inquiries set forth in *Graham v. John Deere Co.*, 383 U.S. 1, 148 USPQ 459 (1966), that are applied for establishing a background for determining obviousness under 35 U.S.C. 103(a) are summarized as follows:

1. Determining the scope and contents of the prior art.
2. Ascertaining the differences between the prior art and the claims at issue.
3. Resolving the level of ordinary skill in the pertinent art.
4. Considering objective evidence present in the application indicating obviousness or nonobviousness.

Claims 1-2, 7, 24-25, and 35 are rejected under 35 U.S.C. 103(a) as being unpatentable over Pighin et al ("Synthesizing Realistic Facial Expressions from Photographs," cited on applicant's IDS) in view of Simon et al (US PGPub 2003/0223622 A1) and Lanitis et al (A. Lanitis, C. J. Taylor, and T. F. Cootes, "Automatic interpretation and coding of face images using flexible models," incorporated by reference into the Simon reference, [0049]).

As to claim 1,

**A computer implemented method for rendering a single frame of a synthesized image, comprising:** (Pighin teaches a computer-implemented method for synthesizing and rendering images – see Figure 4)

**-Generating a geometric component corresponding to a selected image for the frame based on identified feature points from a set of representative images**

(Pighin section 2, page 2, "We...recover the 3D coordinates of a set of feature points on the face..." where this is of a set of representative images, note Figure 4, where 2 exemplary expressions of the actor were captured – Pighin 4 shows generating a geometric component, where clearly teeth and eyes must be generated (section 3.4,

page 5) separately, the base expressions are used to synthesize a final one), **where each image of the set has the identified feature points, and wherein the geometric component is a dimensional vector of feature point positions**; and (Pighin clearly suggests the incorporation of automatic modeling, where the system would find features automatically). Finally Pighin clearly divides facial images into the face, eyes, teeth, and ears (section 3.4, page 5), so the idea of dividing facial images into regions in a set of representative images is clearly taught. Also, Pighin clearly shows (for example, the database of actor / individual expressions capture and shown in Figure 4) that clearly a set of representative exists, and that each image has the same features – they simply move. Clearly the coordinates of those points in 3D would constitute a ‘dimensional vector’)(Simon clearly teaches [0011] that images are segmented into different regions [0049], where a plurality of face images can be used, and that these different regions include skin, eyes, eyebrows, nose, mouth, neck, and hair regions [0011,0044]— see Figure 4 as an example [0050]. Therefore, generating one (1) resultant image would constitute ‘generating a single frame of a synthesized image.’ Clearly, Simon generates new resultant regions based on the results of the image retouching process [0058], where the filters are applied to each sub-image and then the final results are displayed. The system of Simon also generates new features or changed features based on changes in textures [0060]. More importantly, Simon clearly teaches that such regions ‘feature maps’ are generated and/or refined [0080-0082] for the resultant portions of the face. Therefore, each image of the set will have the recited feature points)

**-Generating the selected image for the frame (Pighin Figure 4 as explained above) from a composite of the set of representative images (Pighin Figure 4, representative images – “surprised,” left and “sad,” center – clearly constitute ‘a composite of the set of representative images, since it is generated by a global blend – see caption on Figure 4. Pighin suggests a multi-way blend in section 4.1 on page 6, over the original set of representative images)(Simon blends the new image region portion with the original image region portions – see Figure 13) based on the geometric component (Pighin teaches facial features can be blended based on regional blends in section 4.2, where the mixing proportions for each region varies – see Figure 5 caption and explanation in section 4.2, where clearly a region or sub-portion (e.g. eyes, forehead, nose, and/or the like) would be contemplated)(Simon produces the output image on a per-feature or per-region basis), wherein the selected image and each of the set of representative images comprises a plurality of subregions defined adjacent to each other (Pighin shows that subregions (which correspond to the geometric components) do exist next to each other, since clearly the eyes exist next to the skin portion of the face)(Simon clearly shows in Figure 4 that features such as eyes, nose, eyebrows, hair, etc, exists adjacent to each other – eyebrows are adjacent to eyes, for example), and wherein generating a geometric component is performed for each subregion, (Simon obviously tracks each subregion and generates a new version of it if requested – as explained above [0080-0082], Figure 13, and the like) and wherein generating the selected image comprises generating a composite of the set of representative images based on the corresponding**

**geometric component for each subregion**, (Pighin region blend in section 4.2 and Figure 5, which would (with a multi-way blend, as in section 4.1) generate a composite of the set of representative images)(Simon Figure 13, blending of region of original image and altered or enhanced region of original image, where this clearly represents ) **and rendering a synthesized subregion adjacent to each other with blending at least some boundaries between adjacent subregions without discontinuities in texture in order to generate the selected image.** (Pighin clearly shows how such images are generated adjacent to each other as explained above, where the blending referred to is done with respect to textures – see the suggestion in section 7 (Future work, page 8): “To improve the quality of the composite textures, we could locally warp each **component** texture (and weight) map before blending”. Clearly the idea of blending between adjacent subregions is contemplated or suggested. Pighin synthesizes resultant images as in Figures 4 and 7-8)(Simon very clearly feathers the regional definition masks to create alpha masks [0049]. These feathered binary masks and alpha masks are used in blending operation: “Feathering binary masks and applying the resulting alpha masks in blending operation ensure smooth transitions between regions that have and have not been enhanced. To generate alpha masks the binary masks are feathered by blurring the binary masks with a blurring function where the blur radius is chosen based upon the size of the face ...” Therefore this would clearly constitute ‘synthesizing subregions’ that are adjacent to each other – note previous discussion in first clause, where clearly blending is done at the boundaries via the alpha masks to avoid discontinuities in texture. Indeed, the point of using alpha

masks is such that there will be a continuous texture and there will not be abrupt artifacts that can occur (see Pighin, Future Work, section 7, where it is stated that to improve issues regarding textures ghosting and blurring, local texture warps and blends would improve the situation, which Simon does.)

It would have been obvious to one of ordinary skill in the art at the time the invention was made to combine Pighin and Simon for at least the fact that Simon provides automatic registration and segmentation of images into regions [0011-0012, 0040, and the like], where Pighin suggests adding this feature under 'Automatic modeling' and the fact that Simon provides additional methods of blending local regions together effectively, where Pighin was somewhat silent on how the local blends would take place and how to effectively avoid discontinuities (as noted in the "Improved registration" section under Future Work). It is therefore clear what references teach which limitations.

As to claim 2, Pighin clearly **calculates a plurality of values different from the feature points** by calculating texture maps and weight maps, as explained in sections 3 – 3.2, **wherein a value of the plurality of values is associated with each representative image** since each of the model face images in Pighin has its own texture maps and weight maps, and the **plurality of values are used to composite the set of representative images** where Pighin uses the underlying images to perform global, multi-way, or localized blends – see Figures 4-7 and pages 6-7, and Simon

teaches compositing a set of representative images (e.g. two, the original and the “enhanced” version).

As to claim 7, the feature points in Simon correspond to two-dimensional images, since these can be taken with a single camera and do not involve complex efforts – further Pighin teaches – as in Figure 5 and page 7 – that the user expressions take place as part of a set of photographs.

As to claim 24, examiner submits that Pighin uses a representative face model, where each of the set of representative image is aligned to it (all of page 2, particularly section 2), where that would constitute an underlying reference image (which would be three-dimensional).

Claims 6, 10, 13-14, 26, and 36 are rejected under 35 USC 103(a) as unpatentable over Pighin in view of Simon as applied to claim 1 above, and further in view of Cosatto et al (“Photorealistic Talking Heads from Image Samples”, cited on previous 892)

As to claim 6, examiner submits that Pighin implicitly suggests that one **synthesized subregion is based on a quantity of a set of representatives different than another synthesized region** where Pighin discusses localized blending for expression synthesis in section 4 on page 6.

However, Pighin and Simon do not expressly teach the above. It is submitted that Pighin teaches generating animated transition frames Figure 6 and section 5 on page 7.

Cosatto teaches this limitation (see second paragraph below) and is an analogous art, as explained in this paragraph. In section 1, page 152, that in the first step, image samples of facial parts are generated and results in a database of facial parts. Pages 153-154, section III, teach the methods of how this is done, and how the hierarchy of parts and samples are obtained and subsequently ordered. Section IV on page 154 states that the first step in the process is measuring the face to determine the location of certain facial points (e.g. the recited feature points), which correspond to the "identified feature points" above. The "set of representative images" is the video recorded in section I; all faces would have the same general set of features, e.g. eyes, nose, et cetera. The system of Cosatto clearly synthesizes a geometric component, e.g. synthetic video, with specific emphasis on for example the mouth, section V-B (pages 159-161) with other facial parts discussed in section V-D (page 161), which clearly constitutes "generating a geometric component", and the selected image is simply one frame of video wherein the synthesized face is saying something (e.g. see section V-B).

Since the system of Cosatto is intended primarily for synthesizing the mouth region to create natural-appearing speech, there would obviously be more samples of the mouth region than of other regions, particularly in the database of parts, as that is derived from all the video-recorded phonemes. Therefore, either Cosatto implicitly

teaches it or it is a trivially obvious variant, and it would be obvious to modify for the reasons set forth immediately above, and on page 59, section H, it is stated the mouth database is larger than those of other features and an absolute size provided.

It would have been obvious to one of ordinary skill in the art at the time the invention was made to combine the system of Cosatto with Pighin/Simon such that their system could incorporate pre-recorded speech or video (see for example the suggested Future Work section on pages 8-9 in Pighin) and generate photo-realistic results.

As to claim 10, this is an obvious variant of claim 7, wherein Cosatto in section 2 (pages 152-154, emphasis on page 153) states that three-dimensional images using 3-D scanners are common in the art and in prior work. As further discussed in section IV-A on page 154, feature points on the face are measured in 3-D. Therefore, it would be obvious that the feature points could be on a three-dimensional image and it would be obvious to modify the system of Cosatto to use three-dimensional images for the reasons set forth above. Motivation and rationale is taken from the rejection to claim 6 above.

As to claim 13, Cosatto teaches in section A on page 155 that "Knowing the position of a few points in the face allows to recover the 3-D head pose from 2-D images", where this clearly justifies that examiner's contention that the a few key feature points are used to extract the position of other feature points, see for example section VI-D on page 157. Section V-B on pages 159-160 clearly teaches how knowledge of a few points allows synthesis of a great many essential feature points on the mouth,

which is the key feature. Motivation and rationale is taken from the rejection to claim 6 above.

As to claim 14, Cosatto teaches that obviously feature points are grouped in sets by different regions of the face – see page 154, sections III-1 through III-4 and Fig. 1 or of the synthesized image – see page 161, sections D and E. Finding the position of one feature point on for example the mouth (see section V-B and V-C, particularly page 160) allows the calculation of the shift in other portions of the images, e.g. where the changes in position between one frame and another of the synthesized image are minimized to get more natural appearance (e.g. Figure 7 on page 160), which prima facie tracks change in position of feature points within the mouth region so as to be able to calculate the path that involves the least change in position for Viterbi optimization, and the details on feature point location and tracking are found in sections III-D and III-E, particularly section III-D. Thusly, Cosatto teaches all the limitations. Motivation and rationale is taken from the rejection to claim 6 above.

As to claim 26, this claim is essentially a duplicate of claim 14, with the difference that Cosatto teaches that the feature points are grouped in sets according to the region of the face, e.g. the hierarchical database shown on page 154, and the rest of the limitations are taught in the rejection to claim 14, which is herein incorporated by reference in its entirety. Motivation and combination are taken from claim 6 above.

As to claim 36, this claim is a substantial duplicate of claim 26, with that rejection herein incorporated by reference; motivation and combination is from claim 26 above.

Claims 15-23, 27-29, 33-34, 37-39, and 43 are rejected under 35 U.S.C. 103(a) as being unpatentable over Pighin, Simon, and Cosatto as applied to claim 14 above, and further in view of Chai et al (Chai et al. "Vision-based control of 3D animation".)

As to claim 15, Pighin, Simon, and Cosatto do not expressly teach the limitation of using PCA to track position. Chai teaches the use of PCA on pages 200-201 for example with emphasis on sections 4.2 and 4.3, where it is taught that using PCA, the motion frames are broken down into linear subspaces and motion is tracked in that way. On page 200, section 4.3 it clearly discloses that a database of motion is kept, which would be similar to the database of images in Pighin, Simon, and Cosatto. The database of motion would be with respect to each linear subspace, which obviously could be the different facial regions of Cosatto – that is, the positional changes in motion of the images in the database of Cosatto for facial regions could be found using the PCA techniques of Chai. Therefore, It would have been obvious to one having ordinary skill in the art at the time the invention was made to combine the PCA of Chai with the motion tracking and splitting of the face into different regions of Cosatto and Pighin/Simon for the reasons set forth above, as using PCA allows faster computation times for motion detection and improves temporal coherency (pg. 200- section 4.4 for example).

As to claim 16, it would have been obvious that given that Cosatto tracked motion using overall feature points on the face (e.g. section III-A page 155 or section D page 161) and that Cosatto also tracked feature points within the mouth subset in order

to assure more natural appearing features as the difference between each pose was minimized via Viterbi optimization on page 160, Figure 7. Obviously, overall changes in head position would tracked via the main feature points and determining the necessarily positional changes in the mouth (besides those necessitated by normal motion of talking) would be based on the positional changes in the larger set of feature points on the face itself, e.g. any necessary translational or rotational movement of the overall head for example. Since only the parent references are utilized, no separate motivation or combination is required and that from the rejection to the parent claim is herein incorporated by reference.

As to claim 17, the system of Cosatto has a hierarchical database structure of feature parts, see for example section III, page 154, items 1-3, particularly Item I, titled "Hierarchy of Parts." Since only the Cosatto reference is utilized, no separate motivation or combination is required and that from the rejection to the parent claim is herein incorporated by reference.

As to claim 18, the system of Pighin/Simon and Cosatto does not expressly teach this limitation, insofar as it teaches tracking feature points of the user when the data for the initial sets is recorded, but it does not expressly perform the recited details, although it does monitor feature points of a user. The system of Chai performs the recited limitations, in that it consists of a video camera that monitors the face of a user and generates an image of an avatar making similar facial movements, see for example Fig. 1 on page 193, the caption specifies that users act out the motion in front of a single-view camera, and that the avatars have controlled facial movements similar to those of

the user with texture mapped models (see section 1, left side of page 194, and Figure 2 on page 195, and the captions on it). The system of Chai further tracks feature points of the user (section 2.1, page 196) on the face and moves the avatar as the user moves (see section 1.2 on page 196, where motion data and head motion are separated from facial deformations and then both are applied to the avatar in separate passes).

Obviously, the generated avatars of Chai (Figs. 1 and 2 for example) have separate components of the face, or it would be obvious to use the separate components of Cosatto (+Pighin/Simon) for the face, and to utilize the motion tracking and facial deformation techniques of Chai described above. It would have been obvious to one having ordinary skill in the art at the time the invention was made to combine the systems of Cosatto (and Pighin/Simon) and Chai, since Chai would allow any user to control the facial expressions of an avatar in addition to overlaying audio text and simulating real speech – the facial techniques would allow better synchronization of voice and facial movements in for example the avatars, and would allow even an unskilled user to adequately control facial motions (see section 1, pages 193-194).

As to claim 19, Pighin, Simon, and Cosatto teaches in (Cosatto) section A on page 155 that "Knowing the position of a few points in the face allows to recover the 3-D head pose from 2-D images", where this clearly justifies that examiner's contention that the a few key feature points are used to extract the position of other feature points, see for example section VI-D on page 157. Section V-B on pages 159-160 clearly teaches how knowledge of a few points allows synthesis of a great many essential feature points on the mouth, which is the key feature. Since only the primary references are utilized,

no separate motivation or combination is required and that from the rejection to the parent claim is herein incorporated by reference.

As to claim 20, references Pighin, Simon, and Cosatto do not expressly teach this limitation, insofar as it does teach rendering an image of a speaking human being with the identified feature points on it (see for example Fig. 8, and facial locations are tracked by feature points as illustrated by Fig. 4, where the control points are noted. However, Chai teaches on page 196 in the "initialization" section that the user can select the control points, for which it would be an obvious modification to allow the user to control the movement of a feature point. Also, since the system of Chai (for example, see caption on Fig. 1 on the first page) teaches that the avatar moves in response to user facial and head movements, this also constitutes "receiving information indicative of a user moving a feature point". It would have been obvious to one having ordinary skill in the art at the time the invention was made to combine the systems of Pighin/Simon/Cosatto and Chai, since Chai would allow any user to control the facial expressions of an avatar in addition to overlaying audio text and simulating real speech – the facial techniques would allow better synchronization of voice and facial movements in for example the avatars, and would allow even an unskilled user to adequately control facial motions (see section 1, pages 193-194).

As to claim 21, this claim is a substantial duplicate of claim 16; the rejection to that claim is herein incorporated by reference in its entirety, along with motivation and combination.

As to claims 22 and 23, Pighin, Simon, and Cosatto do not expressly teach this limitation, whilst Chai teaches in Fig. 1 on page 193 that the user can control or select the facial expression by making the desired expression on their own face, e.g. two separate facial expressions are shown in the leftmost column, and in the rightmost column the avatars are shown depicting those facial expressions. Motivation and combination is incorporated by reference from claim 20 above.

As to claim 27, this claim is a substantial duplicate of claim 15, with the only difference that the database of representative images of Cosatto is substituted for the motion database of Chai. Chai teaches a motion database on page 194 on the left side of the page in section 1 and in the caption to Figure 2 on page 195, Chai teaches that the motion database can be used to synthesize expressions. The rest of the limitations are taught in the rejection to claim 15, which are herein incorporated by reference in its entirety; motivation and combination are taken from claim 6 above.

As to claim 28, this claim is a substantial duplicate of claim 16, with that rejection herein incorporated by reference; motivation and combination is from the rejection of claim 27 above.

As to claim 29, this claim is a substantial duplicate of claim 17, with that rejection herein incorporated by reference; motivation and combination is from the rejection of claim 27 above.

As to claim 33, this claim is a substantial duplicate of claim 6, the rejection to which is incorporated herein by reference. Since only the primary reference is utilized,

no separate motivation or combination is required and that from the rejection to the parent claim is herein incorporated by reference.

As to claim 34, Pighin/Simon/Cosatto do not expressly teach this limitation, whilst the system of Chai performs the recited limitations, in that it consists of a video camera that monitors the face of a user and generates an image of an avatar making similar facial movements, see for example Fig. 1 on page 193, the caption specifies that users act out the motion in front of a single-view camera, and that the avatars have controlled facial movements similar to those of the user with texture mapped models (see section 1, left side of page 194, and Figure 2 on page 195, and the captions on it). Motivation and combination are taken from the parent claim, e.g. 25 and herein incorporated by reference.

As to claim 37, this claim is a substantial duplicate of claim 27, with that rejection herein incorporated by reference; motivation and combination is from claim 27 above.

As to claim 38, this claim is a substantial duplicate of claim 28, with that rejection herein incorporated by reference; motivation and combination is from claim 27 above.

As to claim 39, this claim is a substantial duplicate of claim 29, with that rejection herein incorporated by reference; motivation and combination is from claim 27 above.

As to claim 43, this claim is a substantial duplicate of claim 33, with that rejection herein incorporated by reference; motivation and combination is from claim 27 above.

Claims 44 and 47 are rejected under 35 USC 103(a) as unpatentable over Pighin and Simon as applied to claim 1 above, and further in view of Nielsen (US 6,591,011 B1).

As to claim 44, Pighin and Simon do not expressly teach this limitation. Nielsen teaches a method of synthesizing images from a plurality of base images (Abstract), where the system is capable of adjusting tiles or images that have been rotated, transposed, or mirrored to a common frame of reference (Abstract, Figures 14-17B), where image searching and remapping can be done in linear program form using convex hulls (see 18:35-66), where such allow a logarithmic computation cost, which is clearly lower than that of Pighin. It therefore would have been obvious to one of ordinary skill in the art at the time the invention was made to modify Pighin to utilize the convex hull matching method to speed the matching of the images to the underlying model as in page 2, section 2, and the like.

As to claim 47, as noted above in the rejection to claim 44, which is incorporated by reference (and a convex hull is clearly a convex combination of geometry), where clearly Simon and Pighin generate image coefficients, and so does Nielsen. These coefficients are related to the set of representative images and could be typified by e.g. the weight maps of Pighin as discussed above and in section 3. Motivation and combination is taken from the rejection to claim 44 above.

Claim 45 is rejected under 35 USC 103(a) as unpatentable over Pighin and Simon as applied to claim 1 above, and further in view of Stewart et al (US PGPub 2003/0190091 A1).

As to claim 45, Pighin and Simon do not expressly teach this limitation. Stewart teaches the use of an objective function that uses constraints to perform faster image registration to an underlying model and the like [0094], where feature points are used to do so [0044,0122, and the like]. The process and use of objective functions is summarized in [0011-0015]. It would have been obvious to one of ordinary skill in the art at the time the invention was made to modify the system of Pighin to utilize the improved registration methodology of Stewart, since it is faster (for example, [0023-0024]).

Claim 46 is rejected under 35 USC 103(a) as unpatentable over Pighin in view of Simon and Stewart as applied to claim 45 above, and further in view of Fogel et al (US 5,991,459).

As to claim 46, Pighin, Simon, and Stewart do not expressly teach this limitation. Stewart clearly teaches the use of linear programming and linear constraints as explained above, but does not teach that the objective function is a positive semi definite quadratic form with linear constraints. Fogel teaches the use of such functions and such constraints in 25:14-26:50 in the context of image registration between various frames. Therefore, it would have been obvious to one of ordinary skill in the art at the time the invention was made to modify the system of Stewart as above to utilize the

semi-definite quadratic forms and linear constraints of Fogel because it allows the addition of further constraints that shrink the solution space and decrease computational time (24:30-28:50).

Claims 1, 25, and 35 are rejected under 35 U.S.C. 103(a) as being unpatentable over Pighin et al ("Synthesizing Realistic Facial Expressions from Photographs," cited on applicant's IDS) in view of Rowe et al (US PGPub 2003/0202686 A1).

As to claim 1,

**A computer implemented method for rendering a single frame of a synthesized image, comprising:** (Pighin teaches a computer-implemented method for synthesizing and rendering images – see Figure 4)(Rowe Abstract)

**-Generating a geometric component corresponding to a selected image for the frame based on identified feature points from a set of representative images**  
(Pighin section 2, page 2, "We...recover the 3D coordinates of a set of feature points on the face..." where this is of a set of representative images, note Figure 4, where 2 exemplary expressions of the actor were captured – Pighin 4 shows generating a geometric component, where clearly teeth and eyes must be generated (section 3.4, page 5) separately, the base expressions are used to synthesize a final one)(Rowe Fig. 3 shows a face that is segmented into regions using feature points, where these regions would constitute geometric components. Feature points are automatically segmented [0036,0039-0040]. The 3D face model 21 in Figure 1 is created by using a large

number of images of face of individuals are received together with the three-dimensional measurements of the faces appearing in the images [0029-0031]. Next, an average face model is generated from this large number of images [0003,0040,0042,0047]), **where each image of the set has the identified feature points, and wherein the geometric component is a dimensional vector of feature point positions**; and (Pighin clearly suggests the incorporation of automatic modeling, where the system would find features automatically). Finally Pighin clearly divides facial images into the face, eyes, teeth, and ears (section 3.4, page 5), so the idea of dividing facial images into regions in a set of representative images is clearly taught. Also, Pighin clearly shows (for example, the database of actor / individual expressions capture and shown in Figure 4) that clearly a set of representative exists, and that each image has the same features – they simply move. Clearly the coordinates of those points in 3D would constitute a ‘dimensional vector’)(Rowe clearly teaches that this information of component location clearly constitutes a ‘dimensional vector of feature point positions’ since it is indeed a set of representative images. More specifically, the face model 21 is generated as explained in Figure 2 and [0034-0037]. The positions of the feature points are indeed tracked, and clearly an array of such feature points would constitute a dimensional vector. These are used to create the geometric component regions, where clearly Rowe refers to such features as eyes, nose, etc. as facial features [0038].)

**-Generating the selected image for the frame** (Pighin Figure 4 as explained above)(Rowe Fig. 7, step S7-8, [0034] – each mobile phone has its own image

**generation device) from a composite of the set of representative images (Pighin**

**Figure 4, representative images – “surprised,” left and “sad,” center – clearly constitute**

**‘a composite of the set of representative images, since it is generated by a global blend**

**– see caption on Figure 4. Pighin suggests a multi-way blend in section 4.1 on page 6,**

**over the original set of representative images)(Rowe clearly generates the images**

**based on the average face model that is generated from the composite of all images**

**[0003, 0040, 0042, 0047, and the like]) based on the geometric component (Pighin**

**teaches facial features can be blended based on regional blends in section 4.2, where**

**the mixing proportions for each region varies – see Figure 5 caption and explanation in**

**section 4.2, where clearly a region or sub-portion (e.g. eyes, forehead, nose, and/or the**

**like) would be contemplated)(Rowe segments images into geometric components as in**

**Figure 3 and [0040], Abstract, [0009], and the like), wherein the selected image and**

**each of the set of representative images comprises a plurality of subregions**

**defined adjacent to each other (Pighin shows that subregions (which correspond to**

**the geometric components) do exist next to each other, since clearly the eyes exist next**

**to the skin portion of the face)(Simon clearly shows in Figure 4 that features such as**

**eyes, nose, eyebrows, hair, etc, exists adjacent to each other – eyebrows are adjacent**

**to eyes, for example)(Rowe clearly shows that the regions are next to each other as in**

**Figure 3), and wherein generating a geometric component is performed for each**

**subregion, (Rowe clearly at least generating geometric components for each feature –**

**[0029] identifies outlines of facial features and then maps them to the overall average**

**face model and the like, where this would clearly constitute generating a geometric**

component (as for example Figure 4 and the appropriate discussion above) and wherein generating the selected image comprises generating a composite of the set of representative images based on the corresponding geometric component for each subregion, (Pighin region blend in section 4.2 and Figure 5, which would (with a multi-way blend, as in section 4.1) generate a composite of the set of representative images) and rendering a synthesized subregion adjacent to each other with blending at least some boundaries between adjacent subregions without discontinuities in texture in order to generate the selected image. (Pighin clearly shows how such images are generated adjacent to each other as explained above, where the blending referred to is done with respect to textures – see the suggestion in section 7 (Future work, page 8): “To improve the quality of the composite textures, we could locally warp each component texture (and weight) map before blending”. Clearly the idea of blending between adjacent subregions is contemplated or suggested. Pighin synthesizes resultant images as in Figures 4 and 7-8)

It would have been obvious to one of ordinary skill in the art at the time the invention was made to combine Pighin and Rowes for at least the fact that Rowe provides automatic registration and segmentation of images into regions, where Pighin suggests adding this feature under ‘Automatic modeling’ and the fact that Rowes provides additional methods of analyzing facial feature relationships using principal component analysis, where Pighin was somewhat silent on this point but clearly

suggested using PCA in first paragraph on page 9 under 'Future Work'. It is therefore clear what references teach which limitations.

### **Conclusion**

The prior art made of record and not relied upon to make rejections against the independent claims upon is considered pertinent to applicant's independent claims. US 6,591,011 B1 to Nielsen. **Applicant is placed on notice that new grounds of rejection may be added against the independent claims using this reference in the next Office Action.** Applicant is kindly asked to please read this reference, and to please add a brief summary of reasons why the (remaining, amended) independent claims would be patentable with respect to it.

Any inquiry concerning this communication or earlier communications from the examiner should be directed to Eric Woods whose telephone number is 571-272-7775. The examiner can normally be reached on M-F 7:30-5:00.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Ulka Chauhan can be reached on 571-272-7782. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see <http://pair-direct.uspto.gov>. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free). If you would like assistance from a USPTO Customer Service Representative or access to the automated information system, call 800-786-9199 (IN USA OR CANADA) or 571-272-1000.

Eric Woods

September 2, 2006



MARK ZIMMERMAN  
SUPERVISORY PATENT EXAMINER  
TECHNOLOGY CENTER 2600